United Nations
Educational, Scientific and
Cultural Organization

UNESCO Chair on Future Studies
- Anticipation for Sustainability and
Well-being

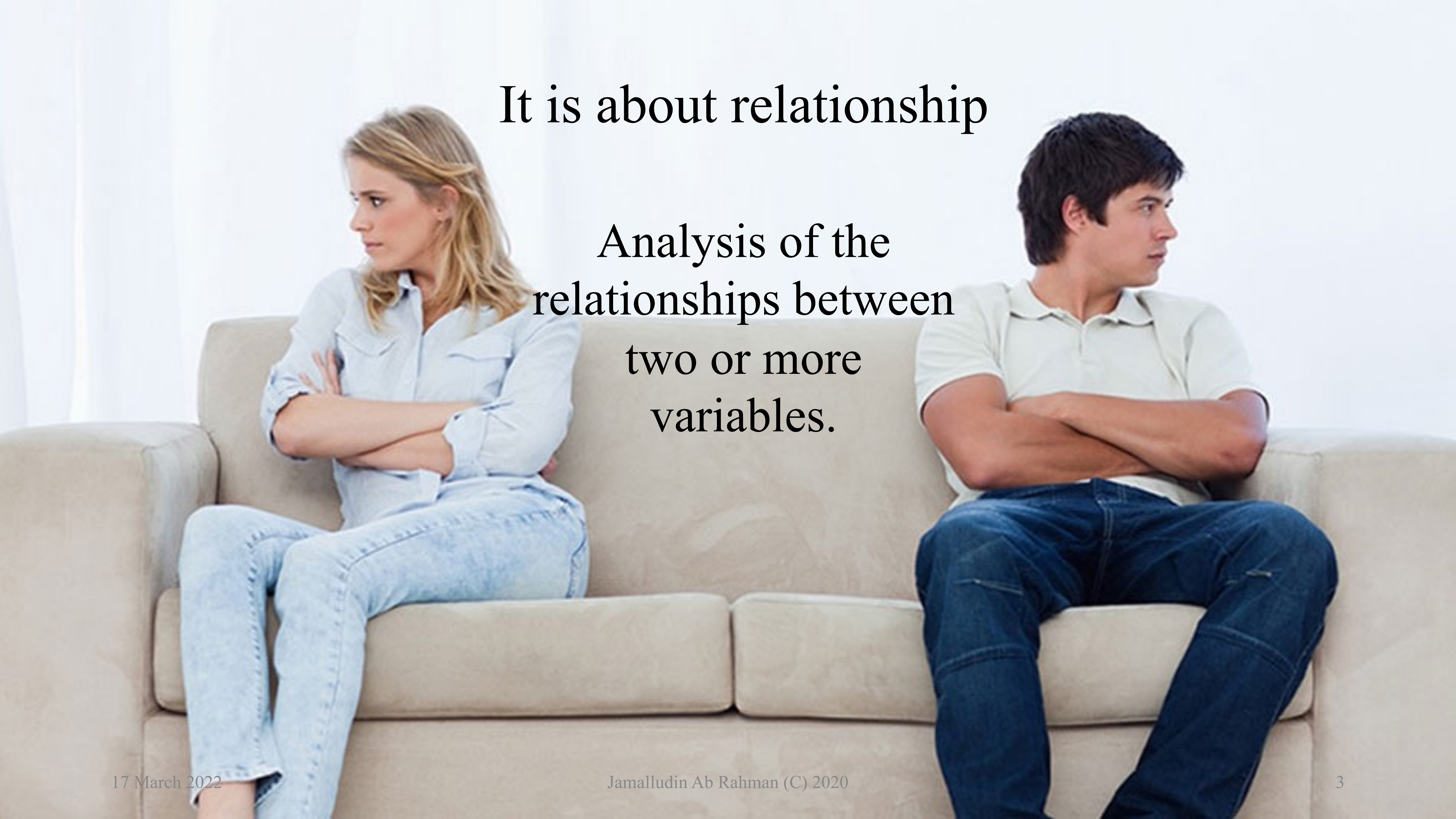
AN INTERNATIONAL AWARD-WINNING INSTITUTION FOR SUSTAINABILITY

Multivariate Analysis

Prof. Dr. Jamalludin Ab Rahman MD MPH
Department of Community Medicine,
Kulliyyah of Medicine

Studying relationship

Jamalludin Ab Rahman



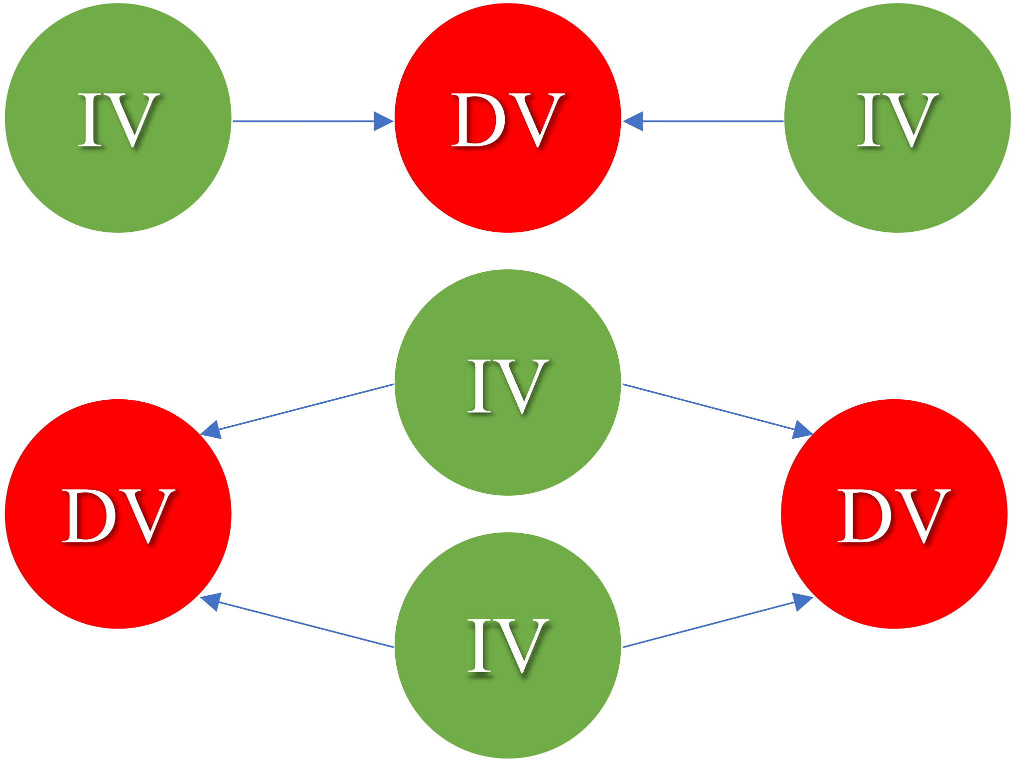
It is about relationship

Analysis of the relationships between two or more variables.

Multivariate?

Basically, bivariate is 1 IV & 1 DV, so multivariate can be:

- More than two IVs
- More than one DV



RESEARCH ARTICLE

Age, sex and disease-specific associations between resting heart rate and cardiovascular mortality in the UK BIOBANK

Zahra Raisi-Estabragh[1,2], Jackie Cooper[1,2], Rebekah Judge[3], Mohammed Y. Khanji[1,2], Patricia B. Munroe[1], Cyrus Cooper[4,5,6], Nicholas C. Harvey[4,5], Steffen E. Petersen[1,2]*

1 William Harvey Research Institute, NIHR Barts Biomedical Research Centre, Queen Mary University of London, London, United Kingdom, **2** Barts Heart Centre, St Bartholomew's Hospital, Barts Health NHS Trust, London, United Kingdom, **3** Imperial College Healthcare Trust, St Mary's Hospital, London, United Kingdom, **4** MRC Lifecourse Epidemiology Unit (MRCLEU), Southampton, United Kingdom, **5** NIHR Southampton Biomedical Research Centre, University of Southampton and University Hospital Southampton NHS Foundation Trust, Southampton, United Kingdom, **6** NIHR Oxford Biomedical Research Centre, University of Oxford, Oxford, United Kingdom

* s.e.petersen@qmul.ac.uk

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7259773/pdf/pone.0233898.pdf

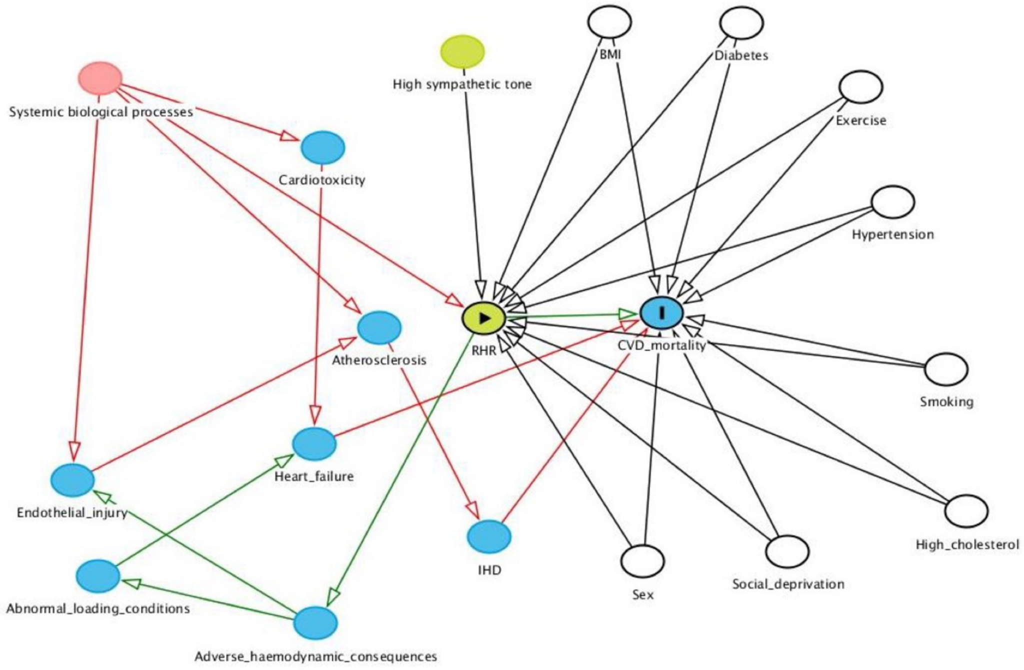


Fig 1. Directed acyclic graph of the potential relationships between cardiovascular disease and resting heart rate. BMI: Body mass index; CVD: Cardiovascular disease; IHD: Ischaemic heart disease; RHR: Resting heart rate. In this model: RHR = exposure and CVD mortality = outcome; Unshaded circles represent true confounders, all controlled for in fully adjusted model.

https://doi.org/10.1371/journal.pone.0233898.g001

Building a Cardiovascular Disease Predictive Model using Structural Equation Model & Fuzzy Cognitive Map

Manpreet Singh, Levi Monteiro Martins, Patrick Joanis and Vijay K. Mago[§]
Lakehead University
Thunder Bay, Ontario P7B 5E1
[§]Corresponding author: vmago@lakeheadu.ca

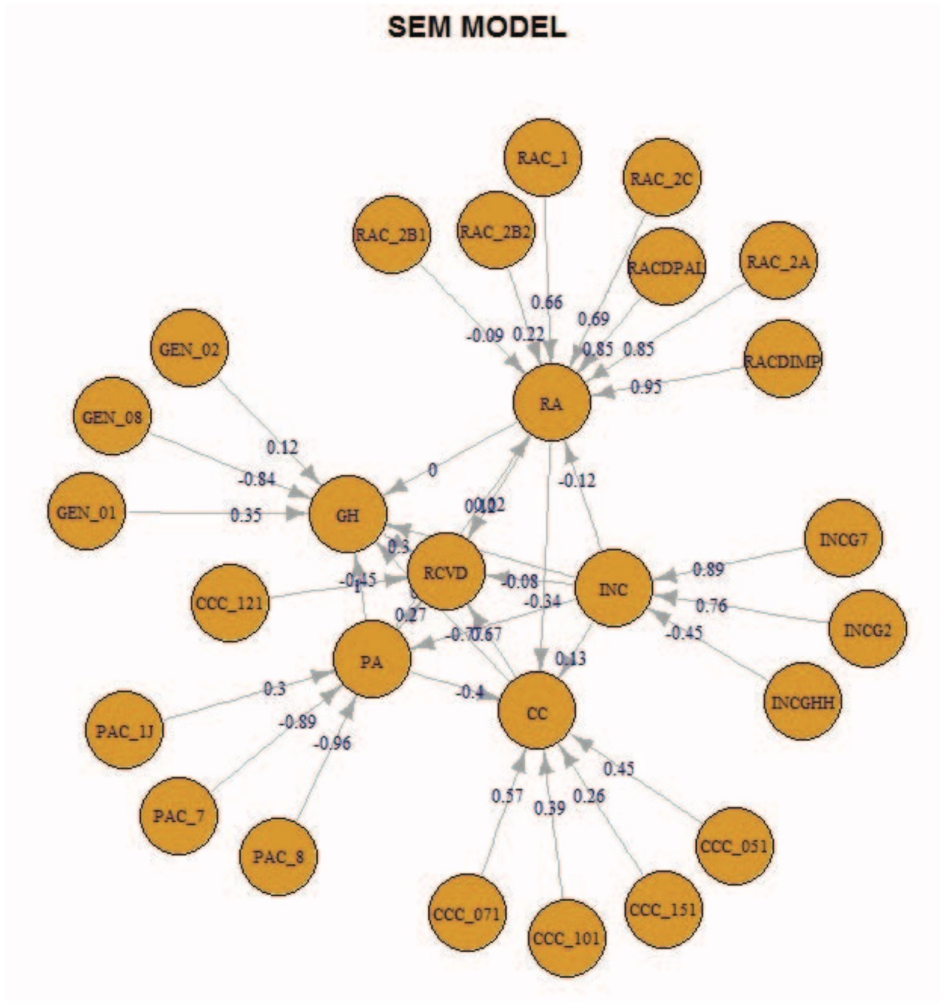


Fig. 2. This SEM model was generated using the 20 variables plus the CCC_121 variable.

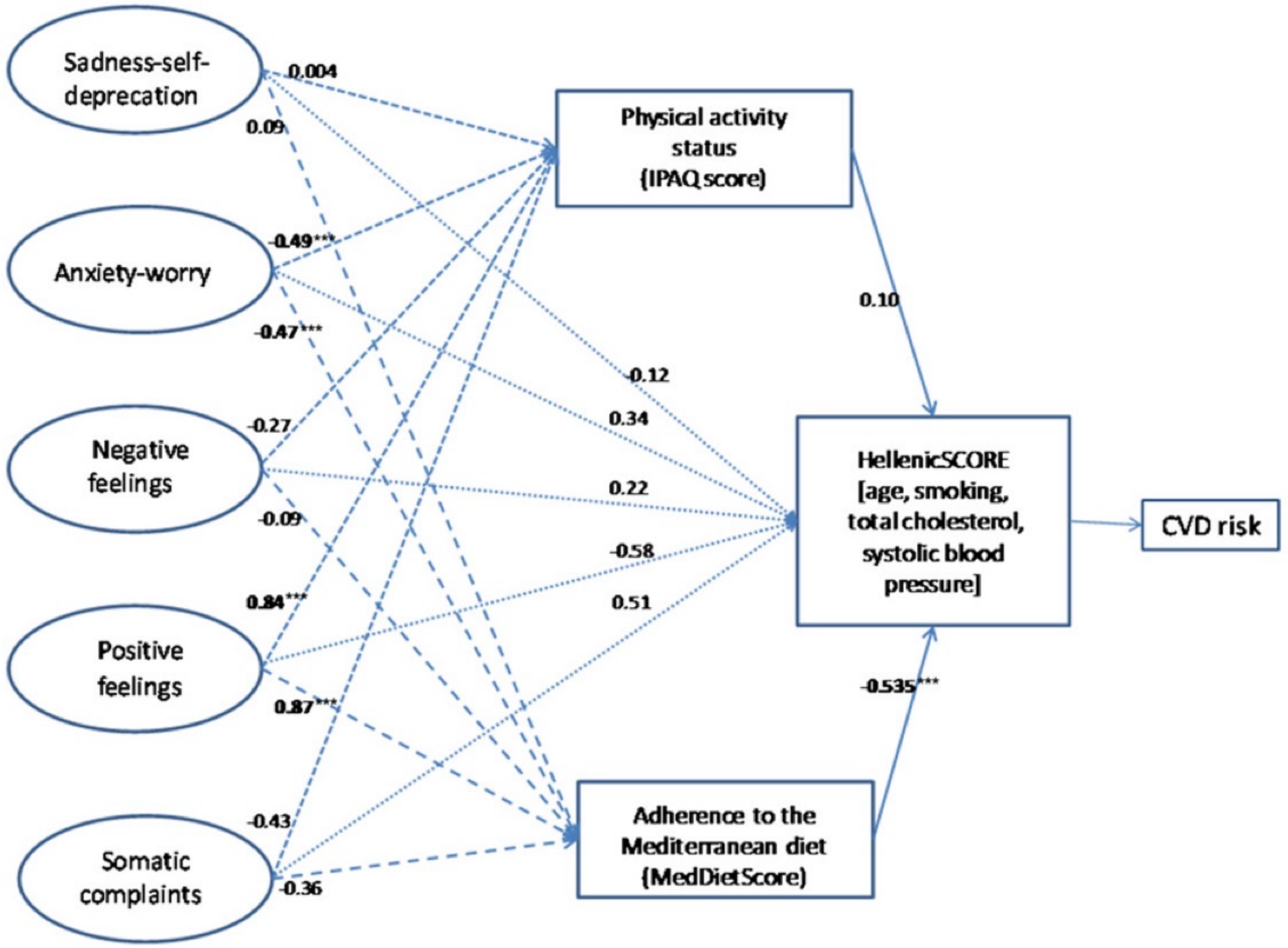


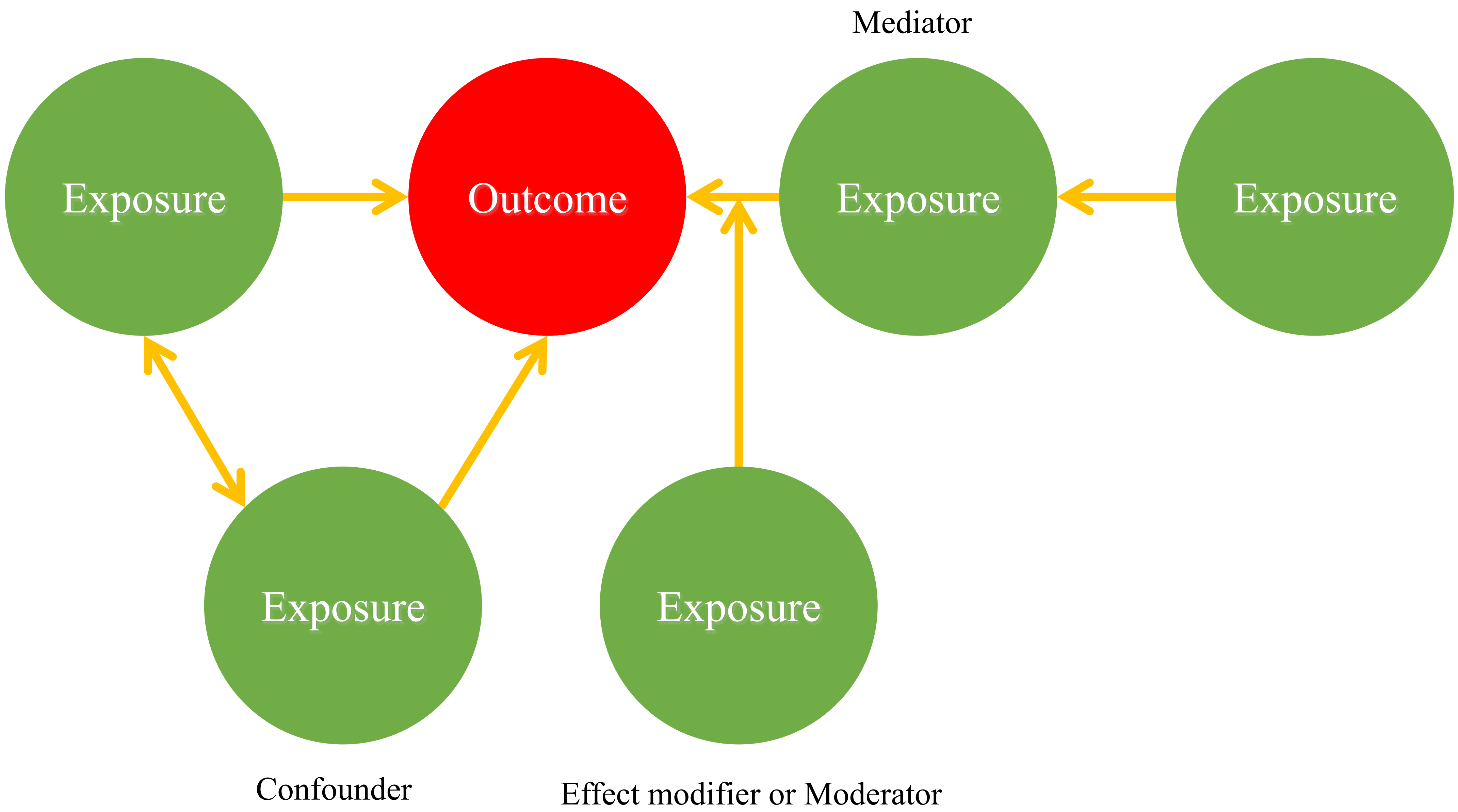
Fig. 1. Conceptual framework derived from the study of depression and anxiety, as well as dietary habits and physical activity status, on CVD risk (measured using the HellenicSCORE), using data from the ATTICA study participants. *Oval shapes represent latent factors, as estimated in the model of anxiety and depression from Table 1. Rectangular shapes represent measured variables: (i) Physical activity (IPAQ score); (ii) adherence to the Mediterranean diet (MedDietScore), (iii) HellenicSCORE (estimated by age, smoking status, total cholesterol, and systolic blood pressure as a proxy for CVD risk estimation). **Numbers represent standardized regression coefficients. The associations between latent and measured variables are represented with dot-line arrows and the associations between measured variables are represented with straight-line arrows. ***$P < .05$.

3^{rd} Factor

Jamalludin Ab Rahman

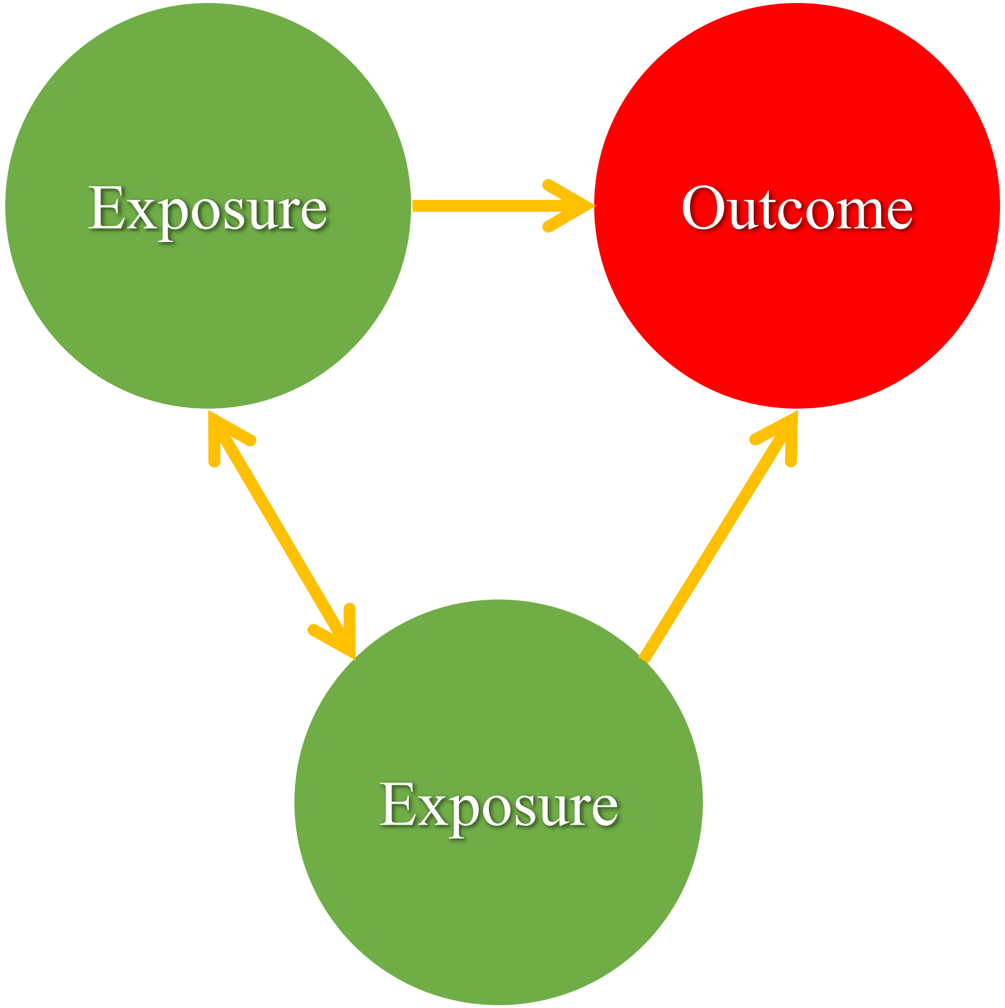
The 3rd factors can be a...

1. Confounder
2. Mediator or intervening factor
3. Moderator or effect modifier (interaction)


Mediator
Exposure
Outcome
Exposure
Exposure
Exposure
Exposure
Confounder
Effect modifier or Moderator

Confounder

Confounder influence a relationship (between two variables), but it is **not a part of the pathway**



Confounder

The unadjusted analysis showed protective effect of current smoking status

Table VI

Unadjusted association between smoking and OA (and severe OA) in knee, hand, foot, and spine (current vs never)

	Grades 2+		Grades 3+	
	RR	CI	RR	CI
Knee	0.62*	0.46–0.83	0.58	0.32–1.04
Hand	0.71*	0.54–0.92	0.38*	0.22–0.67
Foot	0.63*	0.46–0.87	0.95	0.48–1.89
Spine	0.69*	0.54–0.88	0.67	0.43–1.04

*P-value<0.05.
RR=hazard risk ratio; CI=95% confidence interval.

Once adjusted, current smokers has no protective effect on OA anymore

Table VII

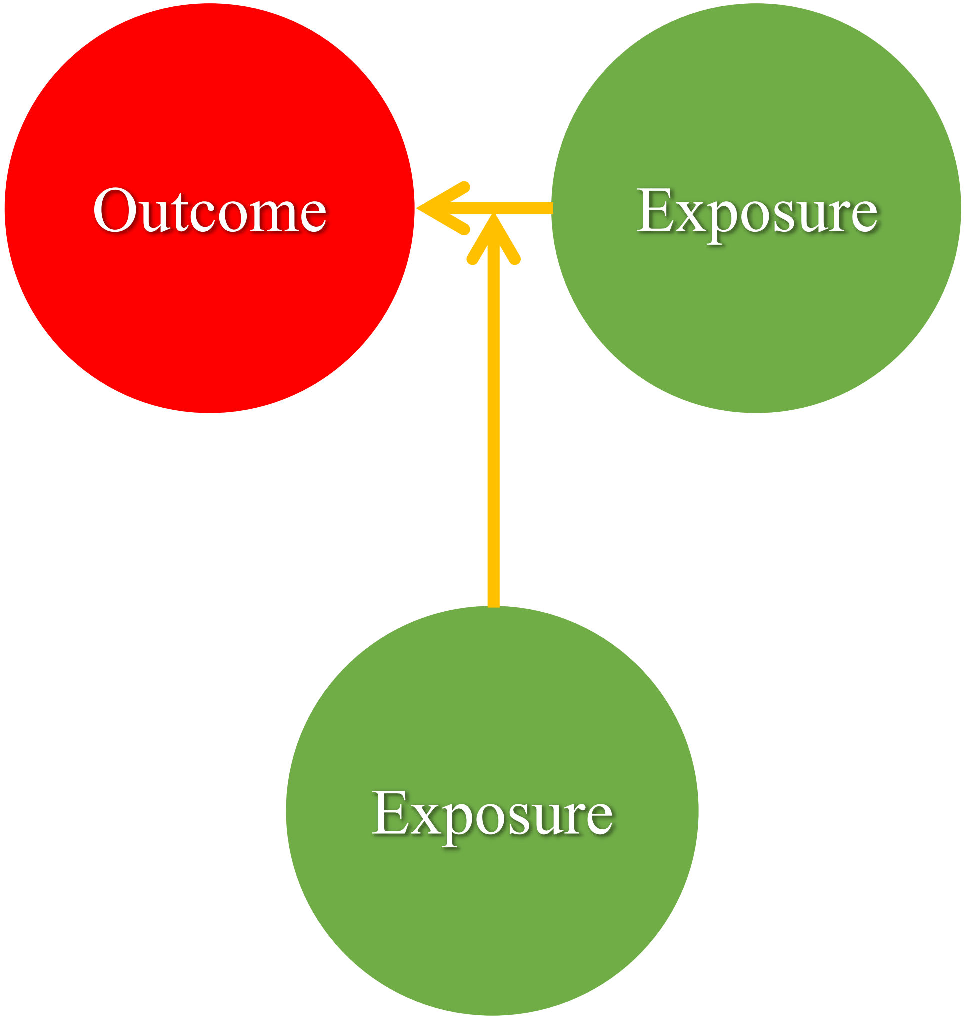
Adjusted association between smoking and OA (and severe OA) in knee, hand, foot, and spine (current vs never)*

	Grades 2+		Grades 3+	
	RR	CI	RR	CI
Knee	0.97	0.71–1.31	1.08	0.58–2.00
Hand	0.99	0.75–1.31	0.60	0.33–1.08
Foot	1.16	0.83–1.63	1.48	0.71–3.08
Spine	0.86	0.67–1.11	0.84	0.53–1.32

*Adjusted for age, body mass index, gender, heredity, occupation, physical activity level, and presence of OA at any of the other three sites.
RR=hazard risk ratio; CI=95% confidence interval.

Moderator

When an exposure has different effects on disease at different values of a variable (**interaction**)



Effect modifier or Moderator

Moderating Effects of Coping on the Relationship Between Stress and the Development of New Brain Lesions in Multiple Sclerosis

DAVID C. MOHR, PHD, DONALD E. GOODKIN, MD, SARAH NELSON, DRRERNAT, DARCY COX, PSYD, AND MICHAEL WEINER, MD

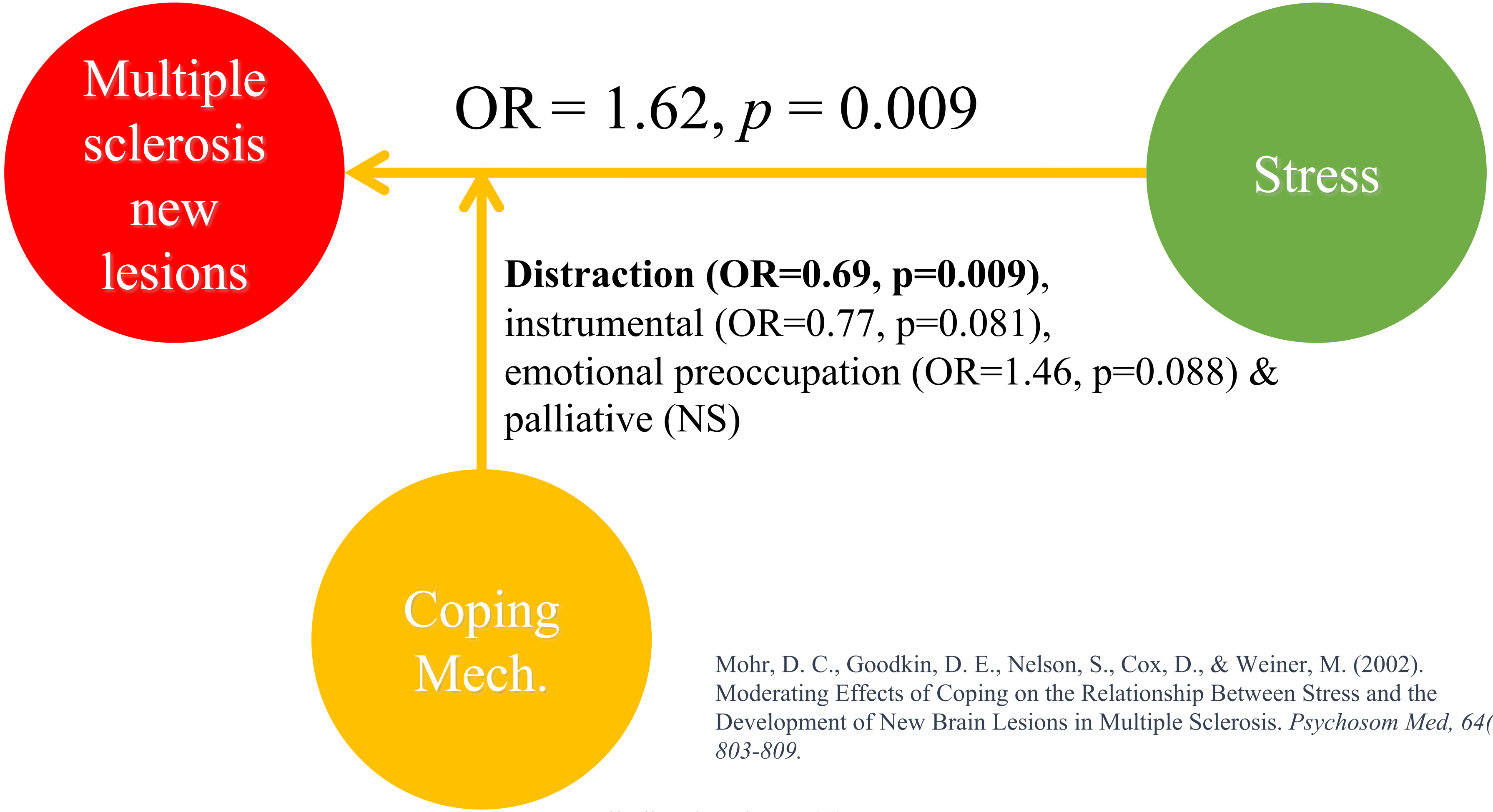
disease exacerbations. Consid-

TABLE 1. Odds Rations for the Effects of Coping and Attachment on the Development of New Gd+ Lesions and Interaction Effects Between Coping or Attachment and Stress and New Gd+ Lesions[a]

Measure	Odds Ratio	Lower 95% CI	Upper 95% CI	*p* Value
New Gd+ lesions at time of stress assessment (8 weeks prior)	**2.55**	**1.38**	**4.70**	**.004**
Treatment with IFNβ-1b	0.50	0.16	1.54	.22
Conflict and Disruption in Routine	**1.62**	**1.12**	**2.34**	**.009**
Coping				
Distraction	1.54	0.78	3.05	.21
Distraction × conflict	**0.69**	**0.49**	**0.98**	**.037**
Palliative	1.14	0.59	2.21	.69
Palliative × conflict	0.83	0.59	1.17	.27
Instrumental	1.26	0.57	2.74	.55
Instrumental × conflict	*0.77*	*0.57*	*1.04*	*.081*
Emotional preoccupation	0.89	0.46	1.72	.72
Emotional preoccupation × conflict	*1.46*	*0.94*	*2.27*	*.088*

[a] Bold = significant relationship; italic = marginal significance.

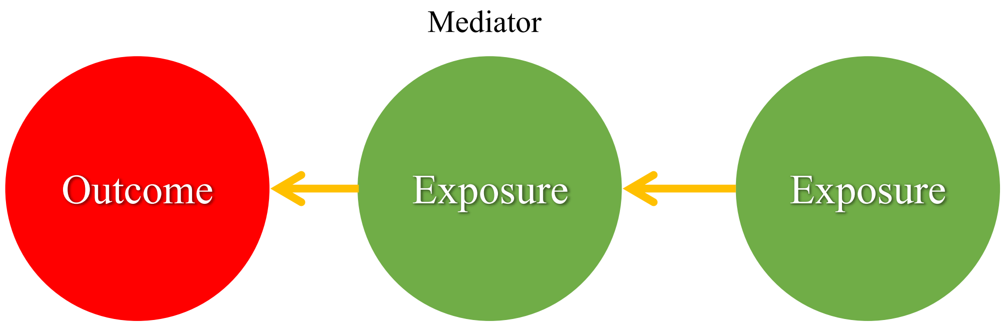
Stress vs. MS vs. Coping mechanism



Mohr, D. C., Goodkin, D. E., Nelson, S., Cox, D., & Weiner, M. (2002). Moderating Effects of Coping on the Relationship Between Stress and the Development of New Brain Lesions in Multiple Sclerosis. *Psychosom Med, 64(5), 803-809.*

Mediator

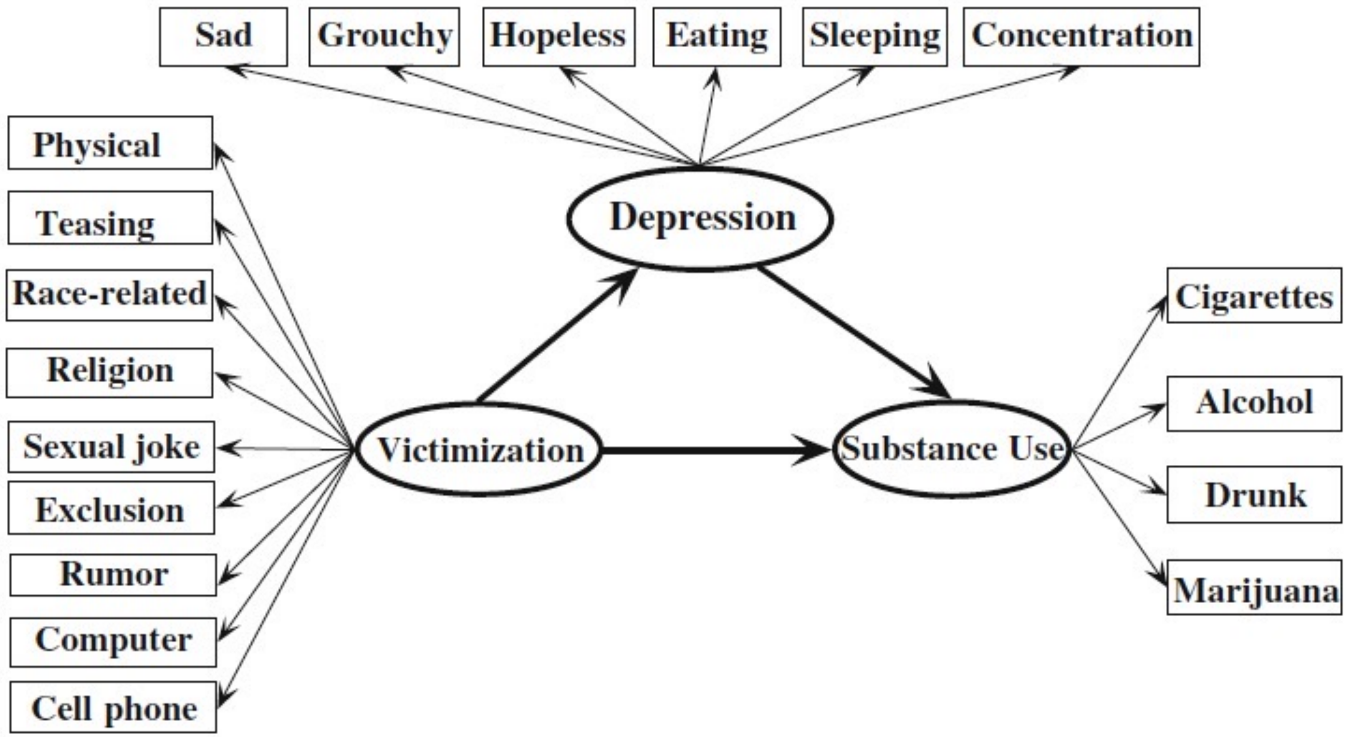
Mediator influence a relationship (between two variables) and it is also **a part of the pathway**



Prev Sci (2010) 11:355–359
DOI 10.1007/s11121-010-0179-0

Bullying Victimization and Substance Use Among U.S. Adolescents: Mediation by Depression

Fig. 1 Theoretical model of the mediating role of depression. This figure shows the hypothesized mediating role of depression on the association of victimization with substance use. All three variables (i.e., victimization, depression and substance use) are latent variables measured by a few indicators. Socio-demographic variables were included as covariates to the outcome variable (i.e., substance use)



poses among adolescent females.

Keywords Victimization · Depression · Substance use · Mediation · Sex difference

online victimization alone. Second, no previous study ... we are aware of has examined potential mediation of the association between victimization and substance use.

Depression as a Candidate Mediator

A body of literature suggests a positive association between ... and depression across various forms of ...

J. W. Luk (✉)
... of Psychology, University of Washington, ... USA

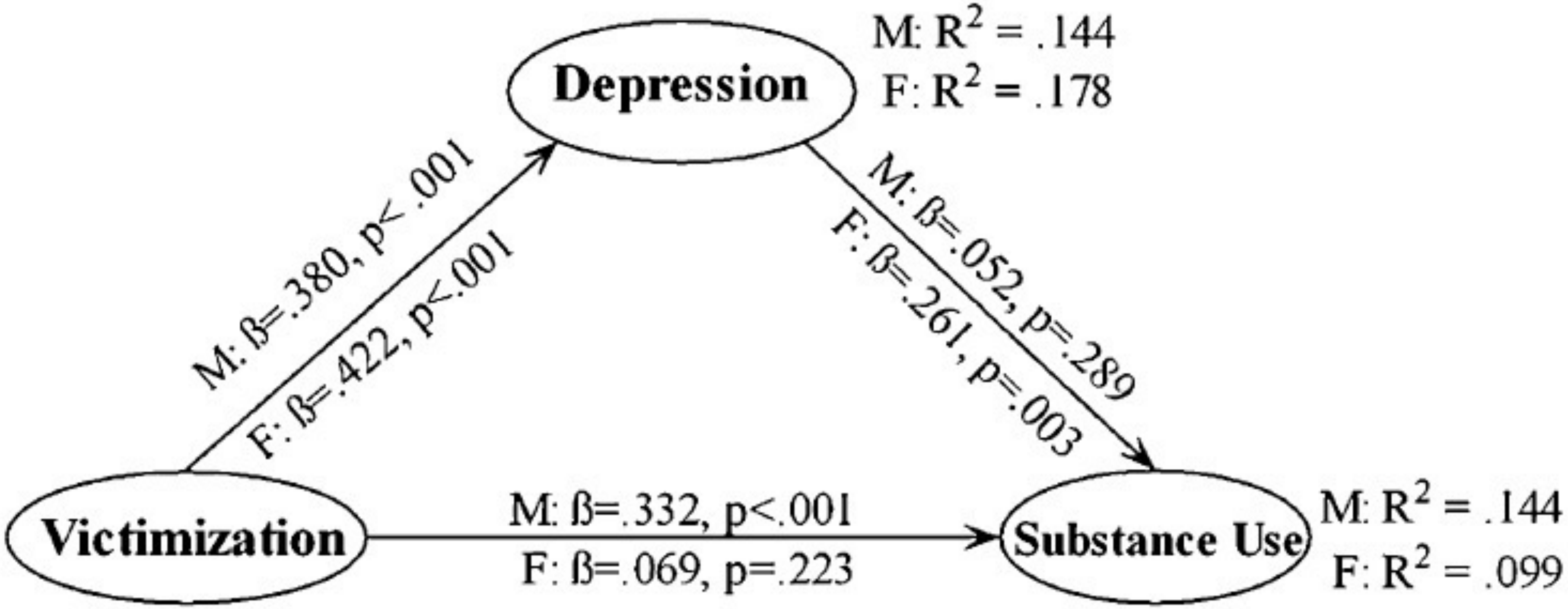


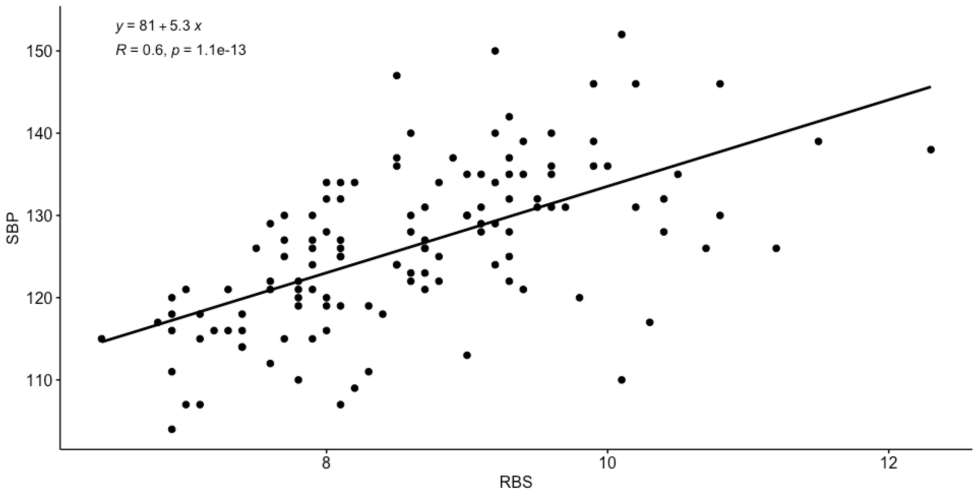
Fig. 2 Results of the mediation model. The multiple-group SEM provided acceptable indices of goodness-of-fit, CFI=.962, TLI=.963, and RMSEA=.048. The path weights in the graph were standardized

Regression

Jamalludin Ab Rahman

Regression

- Best line to fit the data
- Many types:
  - Linear regression (Y = Continuous)
  - Logistic regression (Y= Binary)
  - Poisson regression (Y=Count/Rate)
  - Survival analysis (Y=Time to event)




```
library(ggpubr)
ggscatter(data, x = "RBS", y = "SBP", add = "reg.line") +
  stat_cor(label.x = 6.5, label.y = 150) +
  stat_regline_equation(label.x = 6.5, label.y = 153)
```

Coefficient for Var x_1

Dependent Var

$$Y = \beta o + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_n x_n + \varepsilon$$

Intercept

Error/Residual

Explanatory Var x_1

Linear regression

Example #1

$$Y = \beta o + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_n x_n + \varepsilon$$

Univariate model

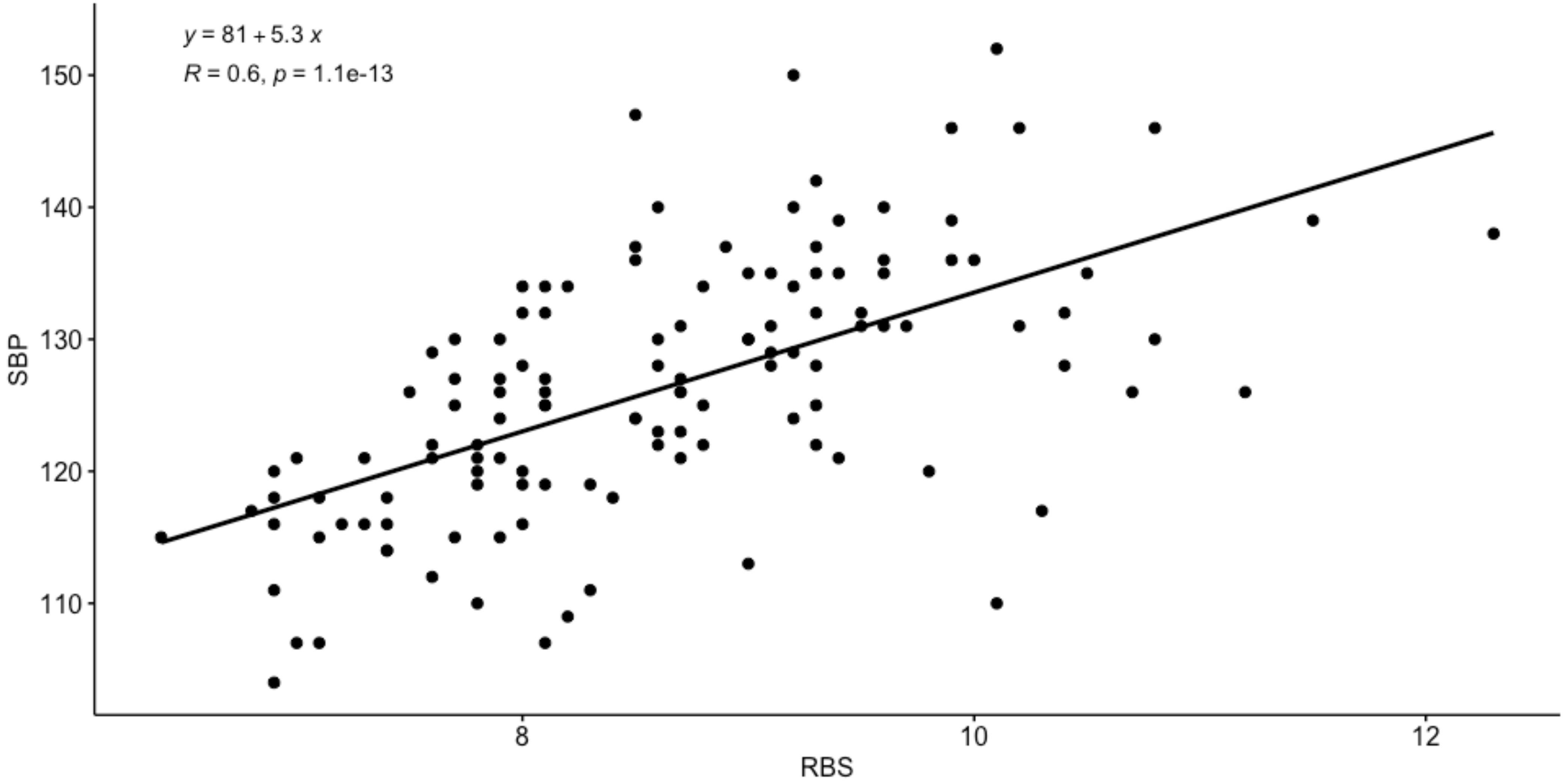
SBP ~ Constant + RBS

Multivariate model

SBP ~ Constant + RBS + Age + Sex + Smoking + Exercise

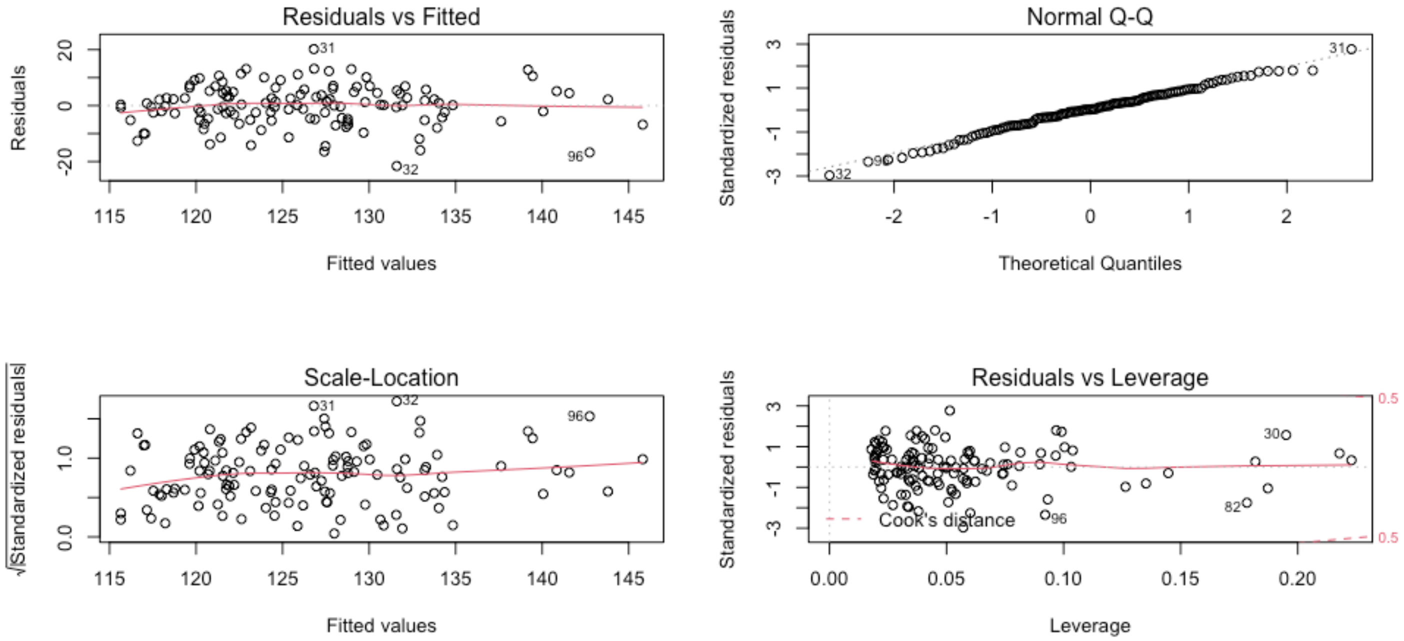
SBP ~ RBS

```
library(ggpubr)
ggscatter(data, x = "RBS", y = "SBP", add = "reg.line") +
  stat_cor(label.x = 6.5, label.y = 150) +
  stat_regline_equation(label.x = 6.5, label.y = 153)
```



Multivariate model

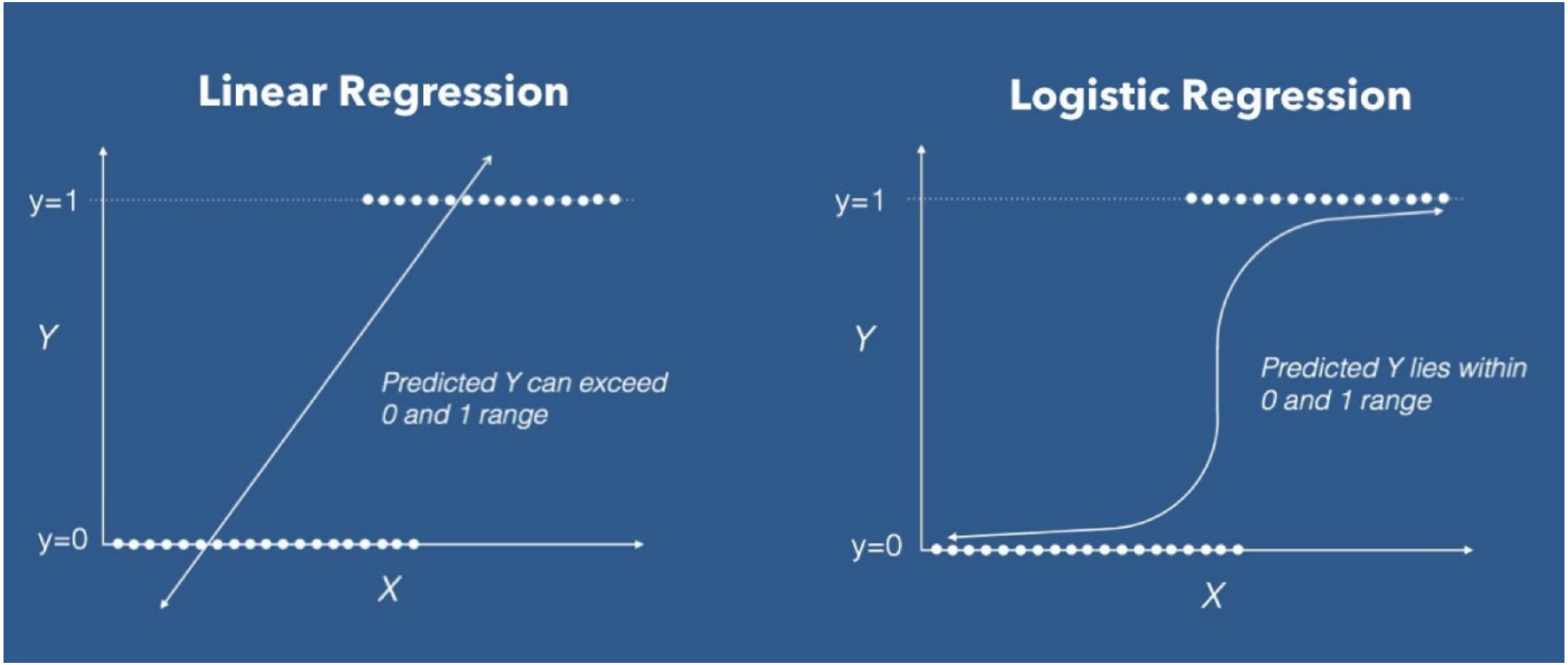
```
> m5 <- lm(SBP ~ Age + Exercise + Sex + Smoking + RBS)
> summary(m5)

Call:
lm(formula = SBP ~ Age + Exercise + Sex + Smoking + RBS)

Residuals:
     Min       1Q   Median       3Q      Max
-21.5948  -4.7332   0.1603   4.8643  20.1945

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  79.5495     6.3796  12.469  < 2e-16 ***
Age           0.1863     0.1012   1.840   0.0683 .
Exercise.L   -5.2211     2.2686  -2.301   0.0231 *
Exercise.Q    2.8326     1.6211   1.747   0.0831 .
SexMale       1.0813     1.4557   0.743   0.4591
SmokingYes    3.6016     1.8133   1.986   0.0493 *
RBS           4.7622     0.6161   7.730 3.66e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.49 on 120 degrees of freedom
Multiple R-squared:  0.4357,     Adjusted R-squared:  0.4075
F-statistic: 15.44 on 6 and 120 DF,  p-value: 4.605e-13
```



Logistic regression

- Odds = $\left(\frac{P}{1-P}\right)$
- $Ln\left(\frac{P}{1-P}\right) = \beta_o + \beta_1 x_1 + \cdots + \beta_n x_n$
- $Ln\left(\frac{P}{1-P}\right)$ = Ln of OR = Logit

Linear vs. Logistic Regression



Source: https://medium.com/@maithilijoshi6/a-comparison-between-linear-and-logistic-regression-8aea40867e2d

Example #2

- Snoring and risk of cardiovascular disease in women . Hu 2000. From The Nurses' Health Study. Cohort. Baseline, N=71,779 women 40 to 65 years old and without diagnosed CVD or cancer in 1986. Till 31st May 1994.
- CVD = Snoring + Age +Smoking + BMI + Alcohol + Physical Activity + Menopausal status + Family history of MI + DM + Cholesterol + Hours sleeping + Sleeping position

Journal of the American College of Cardiology
© 2000 by the American College of Cardiology
Published by Elsevier Science Inc.

Vol. 35, No. 2, 2000
ISSN 0735-1097/00/$20.00
PII S0735-1097(99)00540-9

Risk Factors

Snoring and Risk of Cardiovascular Disease in Women

Frank B. Hu, MD,* Walter C. Willett, MD,*†‡|| JoAnn E. Manson, MD,†‡§|| Graham A. Colditz, MD,†‡|| Eric B. Rimm, ScD,*†|| Frank E. Speizer, MD,‡|| Charles H. Hennekens, MD,¶ Meir J. Stampfer, MD*†‡||

Boston, Massachusetts

OBJECTIVES To examine prospectively the association between snoring and incidence of cardiovascular disease (CVD) in women.

BACKGROUND Whether snoring increases risk of CVD remains unclear; most previous studies have been small, not prospective and limited to men.

METHODS Seventy-one thousand seven hundred seventy-nine female nurses 40 through 65 years of age and without previously diagnosed CVD or cancer at baseline in 1986 were followed up for eight years. Frequency of snoring was assessed using mailed questionnaires at baseline.

RESULTS During eight years of follow-up, we documented 1,042 incident cases of major CVD events (644 coronary heart disease [CHD] and 398 stroke). Compared with nonsnorers, the age-adjusted relative risks (RRs) of CVD were 1.46 (95% confidence interval 1.23 to 1.74) for occasional snorers and 2.02 (1.62 to 2.53) for regular snorers. The age-adjusted RRs of CHD were 1.43 (1.15 to 1.77) for occasional snorers and 2.18 (1.65 to 2.87) for regular snorers. The age-adjusted RRs of stroke were 1.60 (1.21 to 2.12) and 1.88 (1.29 to 2.74), respectively. After further adjustment for smoking, body mass index (BMI) and other covariates, the positive association between snoring and CVD was attenuated but remained statistically significant (RRs of CVD were 1.20 [1.01 to 1.43] for occasional snorers and 1.33 [1.06–1.67] for regular snorers.

CONCLUSIONS These data suggested that snoring is associated with a modest but significantly increased risk of CVD in women, independent of age, smoking, BMI and other cardiovascular risk factors. While further study is needed to elucidate the biological mechanism underlying this association, snoring may help clinicians identify individuals at higher risk for CVD. (J Am Coll Cardiol 2000;35:308–13) © 2000 by the American College of Cardiology

Table 2. Relative Risks (95% Confidence Intervals) of Coronary Heart Disease and Stroke According to Self-Reported Snoring at Baseline in 1986

	Frequency of Snoring		
	Never	Occasionally	Regularly
Total cardiovascular events (coronary heart disease + stroke)**			
# cases	162	729	151
Person-years	143,719	356,530	51,292
Age-adjusted	1.0	1.46 (1.23–1.74)	2.02 (1.62–2.53)
Multivariate	1.0	1.20 (1.01–1.43)	1.33 (1.06–1.67)
Hemorrhagic stroke			
# cases	16	79	13
Age-adjusted	1.0	1.77 (1.03–3.03)	1.97 (0.94–4.11)
Multivariate	1.0	1.71 (0.99–2.95)	1.80 (0.85–3.83)

*Models include the following: age (five-year category); time period (four periods); body mass index (five categories); cigarette smoking (never, past and current smoking of 1 to 14, 15 to 24 and ≥25 cigarettes per day); menopausal status (premenopausal, postmenopausal without hormone replacement, postmenopausal with past hormone replacement, postmenopausal with current hormone replacement); parental history of myocardial infarction before 60 years of age; alcohol consumption (four categories); multivitamin and vitamin E supplement use; physical activity (METS, in quintiles); average number hours of sleeping (5 h, 6, 7, 8, ≥9); usual sleep positions (on side, on back, on front); history of diabetes; history of hypercholesterolemia.
METs = metabolic equivalent hours.

Table 2. Relative Risks (95% Confidence Intervals) of Coronary Heart Disease and Stroke According to Self-Reported Snoring at Baseline in 1986

	Frequency of Snoring		
	Never	Occasionally	Regularly
Total cardiovascular events (coronary heart disease + stroke)**			
# cases	162	729	151
Person-years	143,719	356,530	51,292
Age-adjusted	1.0	1.46 (1.23–1.74)	2.02 (1.62–2.53)
Multivariate	1.0	1.20 (1.01–1.43)	1.33 (1.06–1.67)

The steps

Jamalludin Ab Rahman

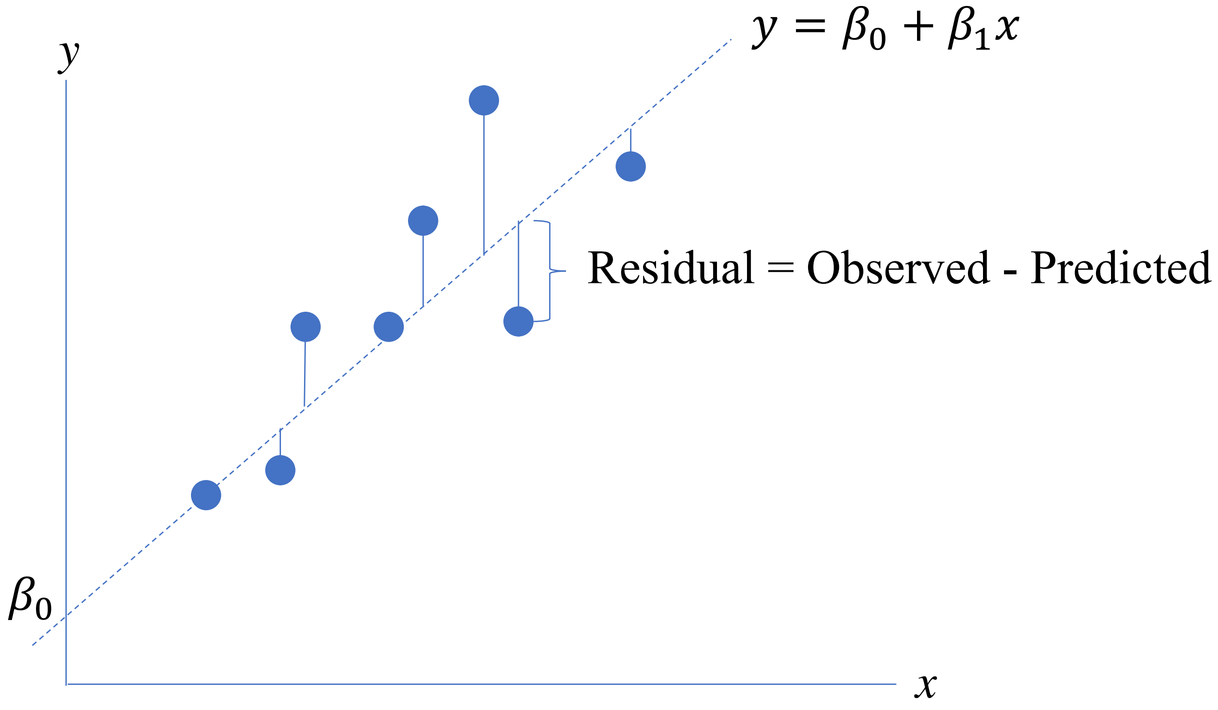
The steps

1. Building a model
2. Interpreting coefficients
3. Assessing 3rd factors
4. Examining assumptions
5. Describe, assessing predictive & effect size

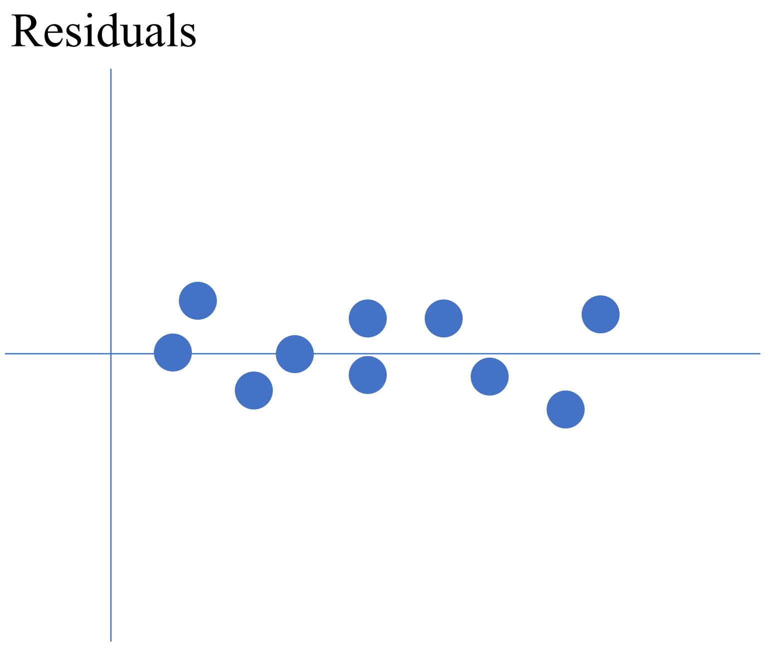
Testing assumptions for linear regression

1. **Linearity** between DV & IVs – *Scatter plot residuals vs. predicted*
2. **Normality** – *Histogram of residuals*
3. No outliers – *Casewise diagnostics (within +/- 2SD), Cook's D (for influential points, <1), leverage point (< 4/n)*
4. **Homogeneity** – Scatter plot
5. **Independence** (no autocorrelation) – *Durbin Watson 1.5-2.5 (or some says 0-4)*
6. **No multicollinearity** – *Tolerance > 0.1, VIF <10*

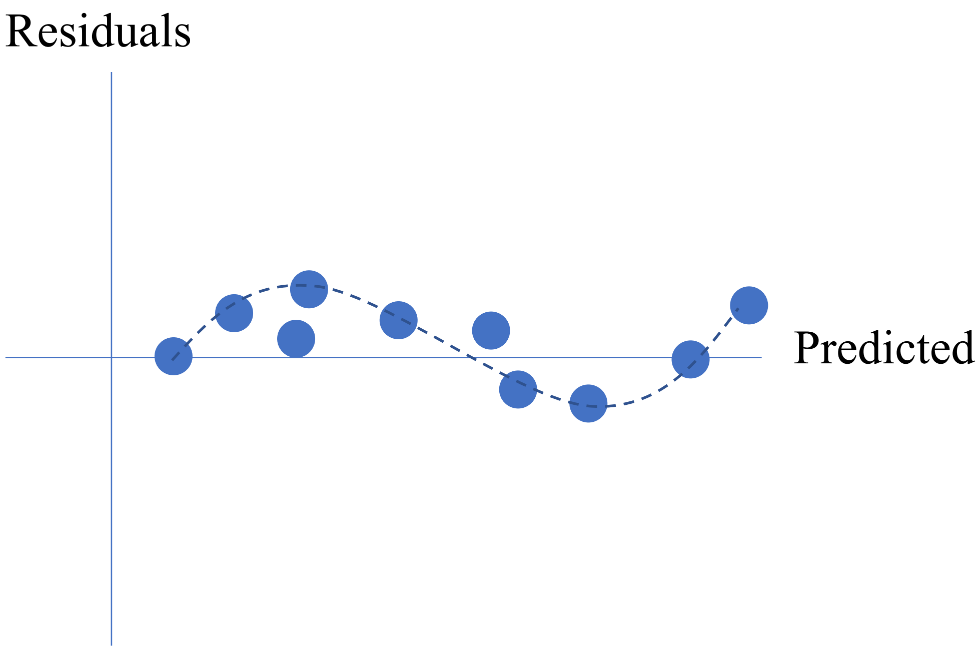
Residuals



Residual statistics - Linearity

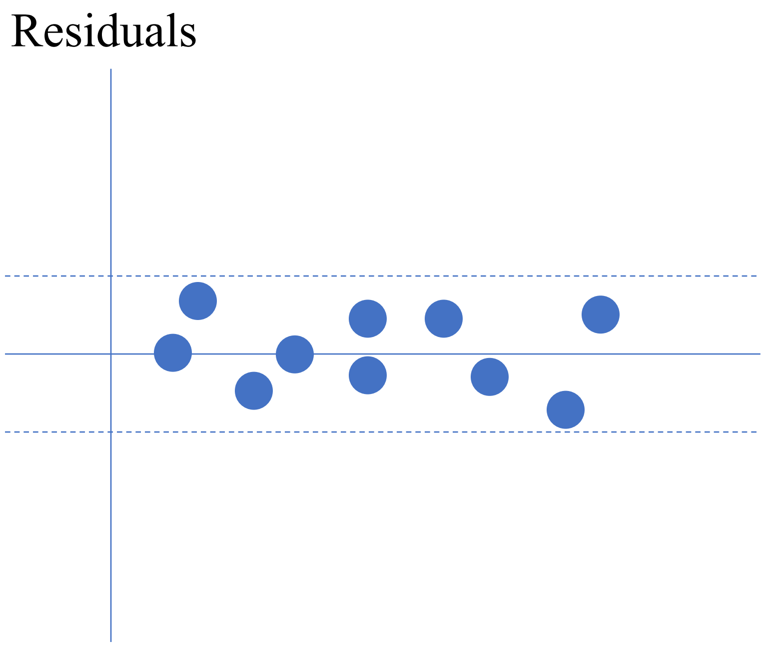


Linear

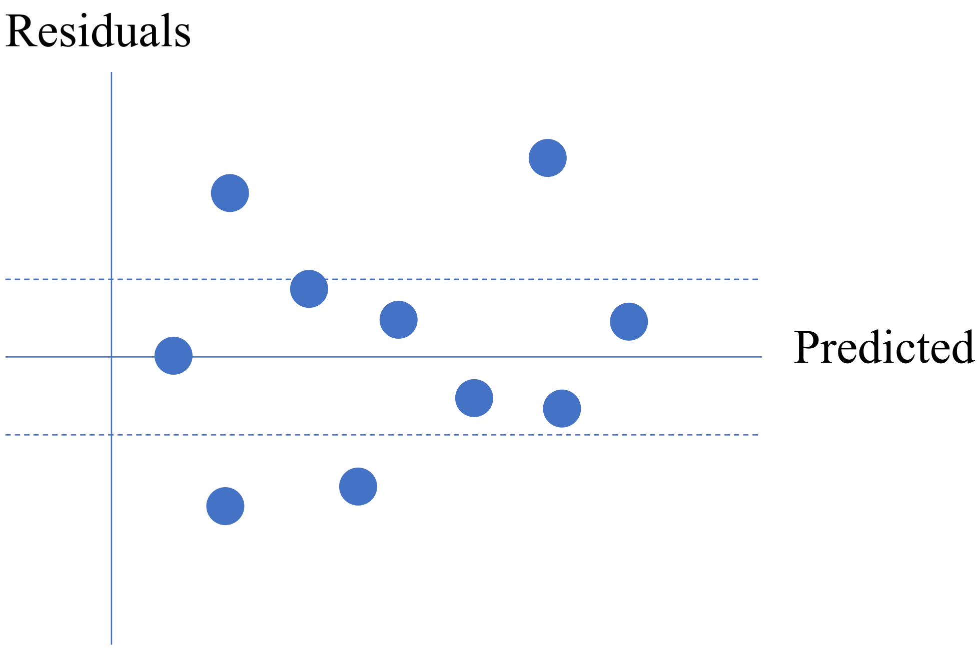


Non linear

Residual statistics - Normality

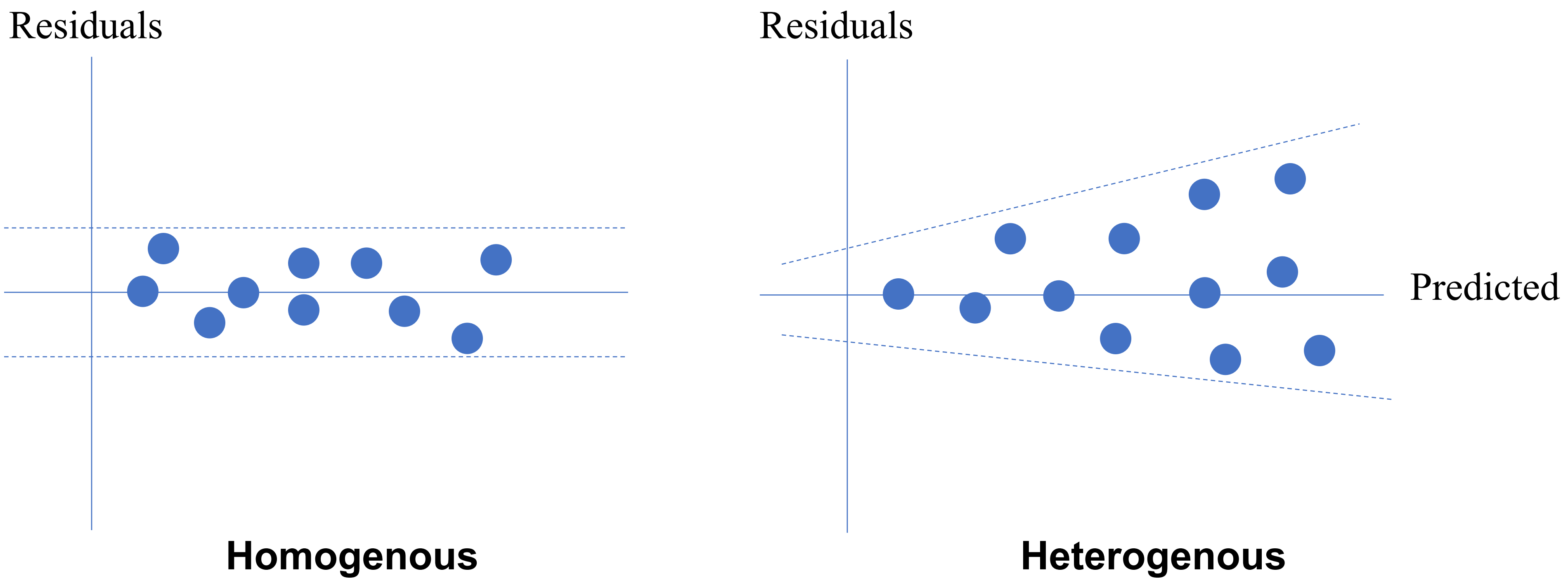


Normal distribution

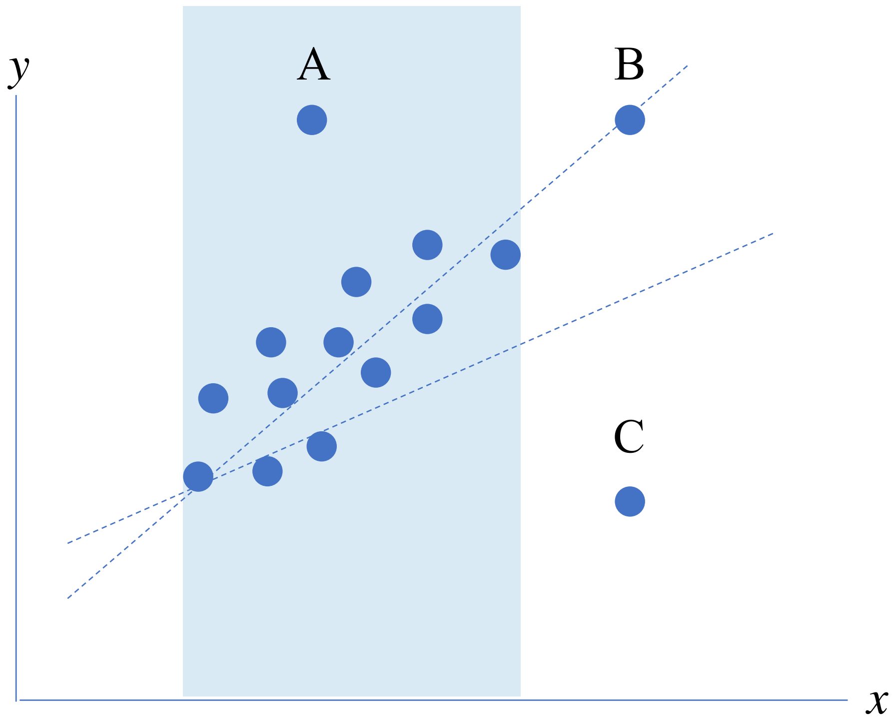


Not Normal

Residual statistics - Heterogeneity



Influential data

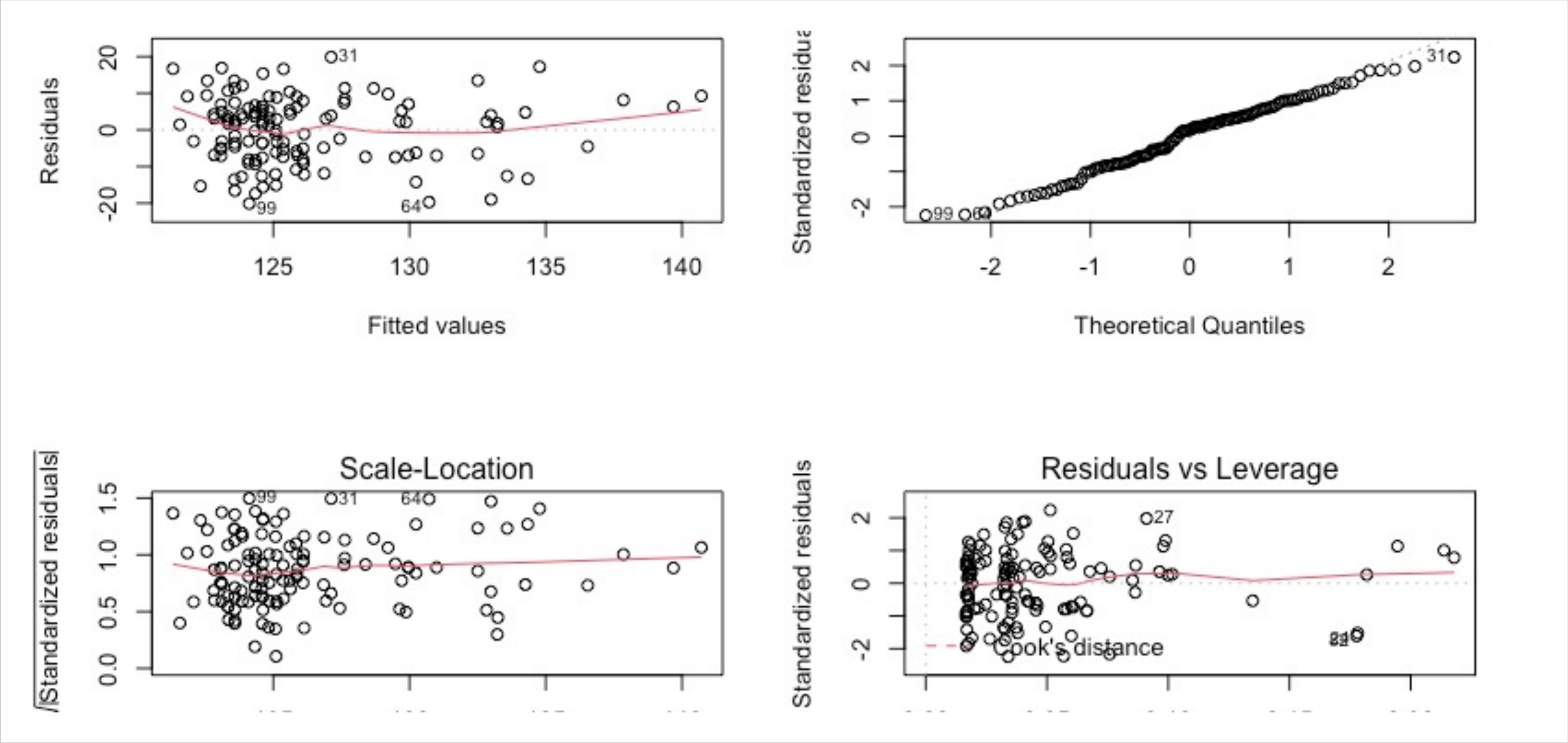


A = **Outlier**, still within the range of *x*, large residual value

B & C = **Leverage** points

B = Good leverage, it won't impact the regression line

C = Bad leverage. It will change the regression line



Checking assumptions

Type of multivariate tests

Dependent Variables	Independent Variables	Test
1 – Cont	All Cont	Linear Regression
1 – Cont	All Cat	ANOVA
1 – Cont	Cont + Cat	ANCOVA
> 1 – Cont	All Cat	MANOVA
> 1 – Cont	Cat + Cont	MANCOVA
Repeated Cont	Cat + Cont	Repeated measures ANOVA
1 – Dichotomous	Cont + Cat	Binary Logistic Regression
1 - Polytomous	Cont + Cat	Multinomial Logistic Regression